Transitioning PDS from RMI to HTTP

Sean Kelly
NASA Jet Propulsion Laboratory
California Institute of Technology

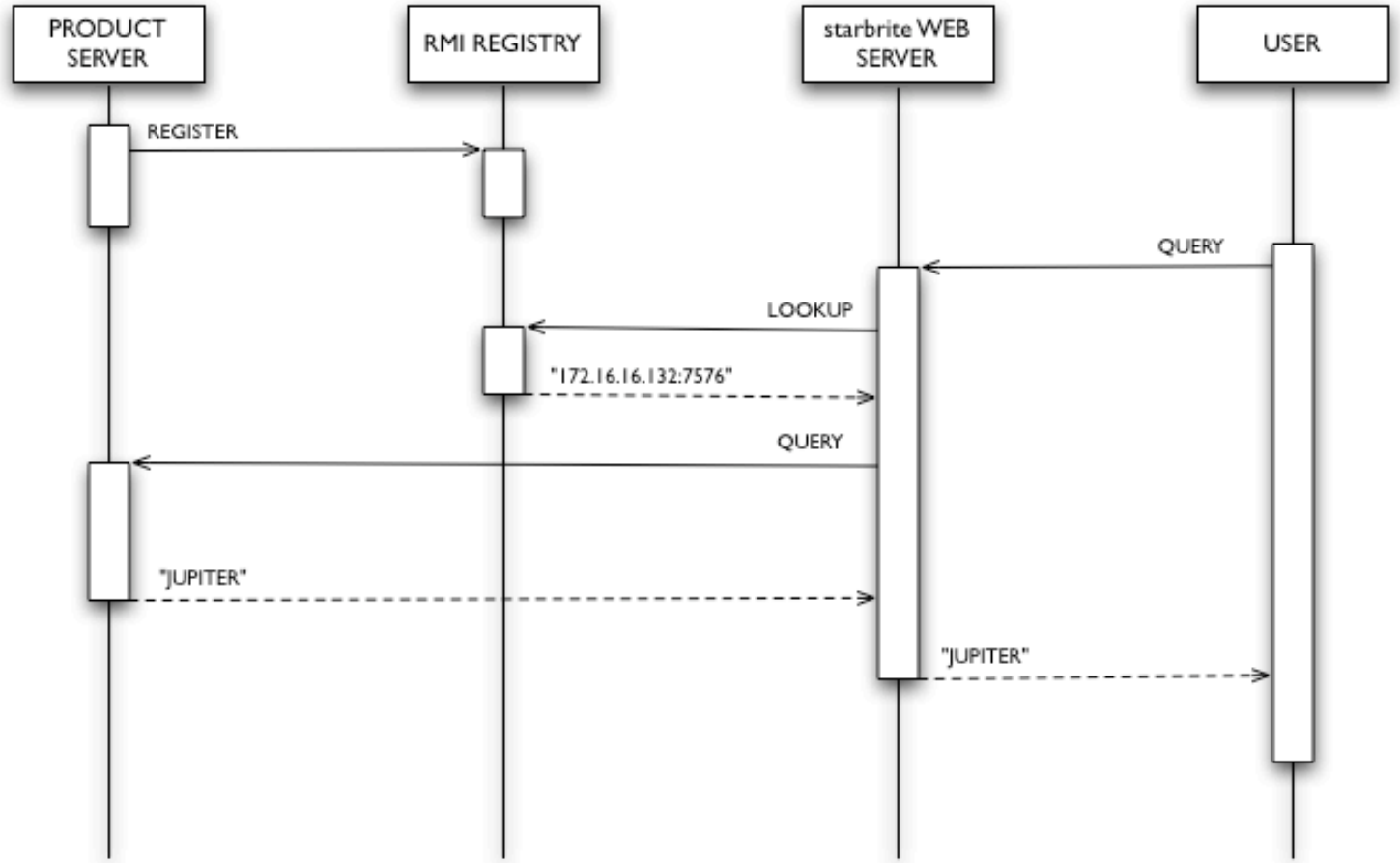
Overview

- Current network implementation
  - Its disadvantages
- Using HTTP instead
  - Its advantages
- Upgrade packages
- Performance

Current Network Technology

- Uses Java's Remote Method Invocation
- Product servers are treated as remote Java objects:
  - They *register* with a naming registry
  - They're *queried* by the starbrite web app
  - They *respond* by serializing blocks of data

System Architecture



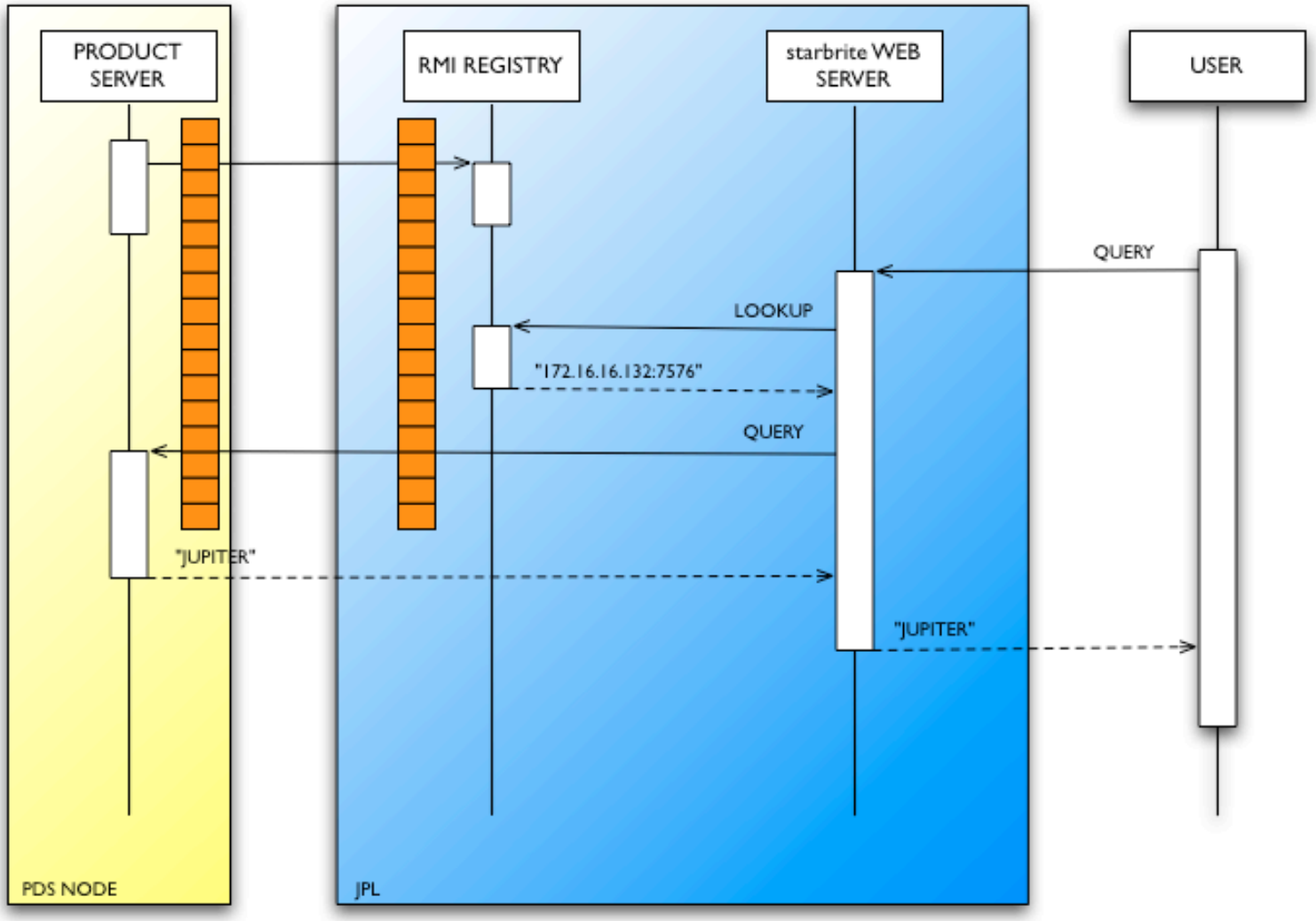
RMI Drawbacks

- Naming registry is a point of failure
  - Even if servers are up and running, clients cannot locate them
  - PDS servers have moved only *twice* in 3 years
- Slower
  - Cannot stream data over RMI
  - Must send blocks of data, one block at a time

RMI Drawbacks

- Naming registry requires changes to the security policy at *multiple institutions*
  - Servers must make *outgoing* calls to registry
  - JPL must accept those incoming registrations

Firewalls


PRODUCT SERVER
RMI REGISTRY
starbrite WEB SERVER
USER
QUERY
LOOKUP
"172.16.16.132:7576"
QUERY
"JUPITER"
"JUPITER"
PDS NODE
JPL

Firewalls Change Constantly

- Admins update firewall rules all the time
  - Typically, they remove exceptions
- Suddenly preventing outgoing registrations
  - Or preventing incoming queries

Using HTTP Instead

- No central naming registry
  - We need to know your service endpoints
    - But they *rarely* change
- Streaming profiles and products
  - As fast as HTTP can be
- Compatible
  - On both server and client side
  - **No interface changes at all**

Current Interfaces Unaffected

- URLs to favorite products are the same: `http://starbrite.jpl.nasa.gov/prod?object=...`
- Existing product and profile handlers are the same
  - Node-developed handlers will still work
  - No code modifications necessary

Better Reliability

- No longer hindered by outgoing registrations
  - And can remove those firewall exceptions
- Many institutes happily pass HTTP
  - Server can use port 80
    - Or any other port
    - Default is same as old: 7576

Deployment

- We'll upgrade "starbrite"
  - New software will work with *both* RMI and HTTP product servers
- Then each node may upgrade at its leisure
  - Continue to run RMI until you get around to upgrading

Upgrade Packages

- Pre-made installers for instant setups:
  - http://oodt.jpl.nasa.gov/pds-product-servers/
  - Installers for Mac OS X, Windows, Linux (RPM), and Generic Unix
- Includes:
  - PDS query handler
  - HTTP product server and OODT code
  - A lightweight web server

Or Roll Your Own

- If you already have a web server you want to use:
  - Make sure it's compatible with 2.3 of the servlet spec
  - Download from `http://oodt.jpl.nasa.gov/pds-web-ps/` instead
  - Deploy the war file using your server's tools
- Tested web servers
  - Tomcat 5.5.9 (not compatible with 5.0 or 4)
  - Jetty 5.1.4

Performance

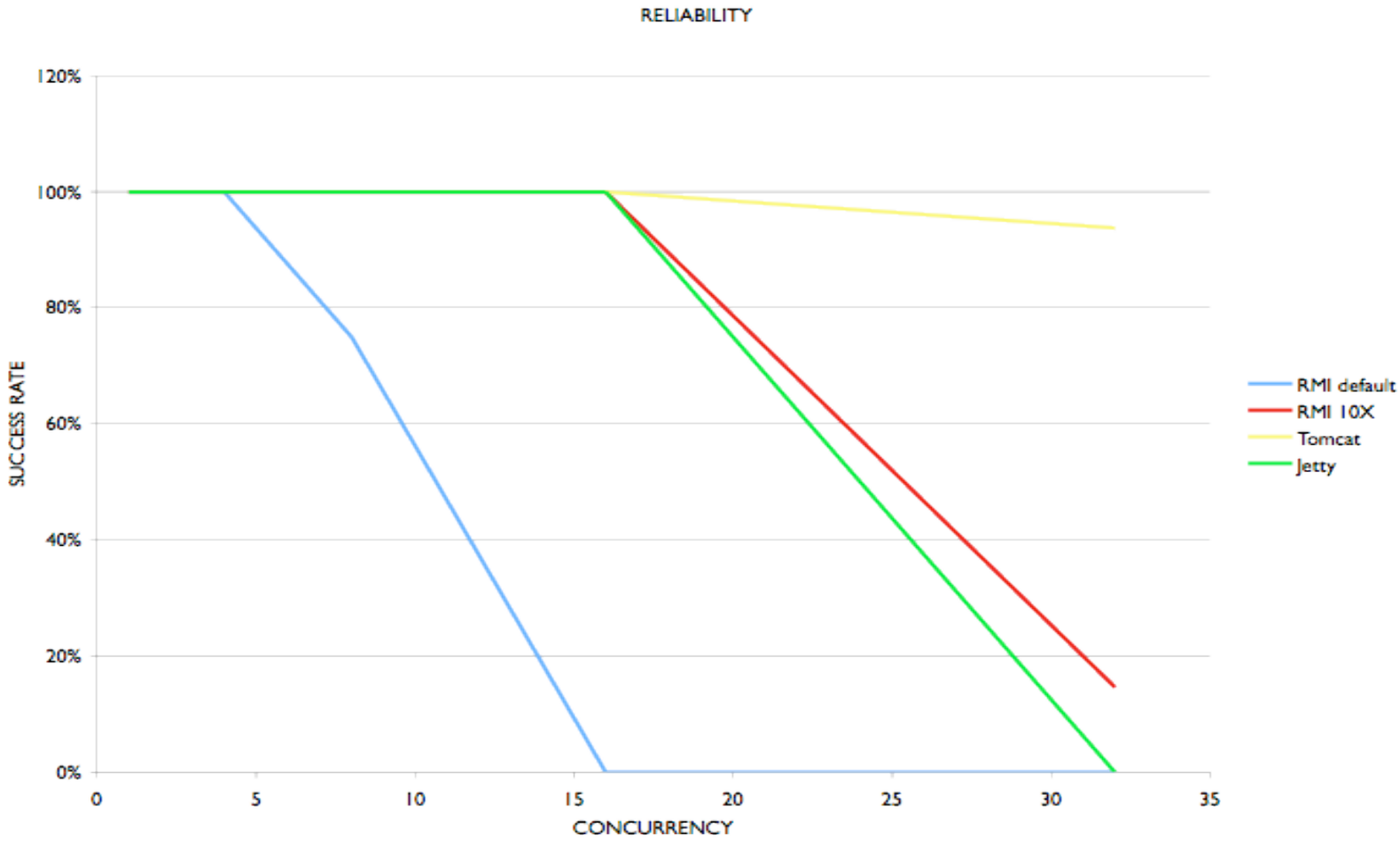
- System configuration
  - Server: Power Mac 2.0GHz dual G5, Mac OS X
  - Client: Mac Mini 1.42GHz G4, Mac OS X
  - 100baseTX fast Ethernet
- Tested configurations
  - RMI (default block size, 4096 bytes)
  - RMI (10X block size, 40960 bytes)
  - HTTP under Tomcat 5.5.9
  - HTTP under Jetty 5.1.4

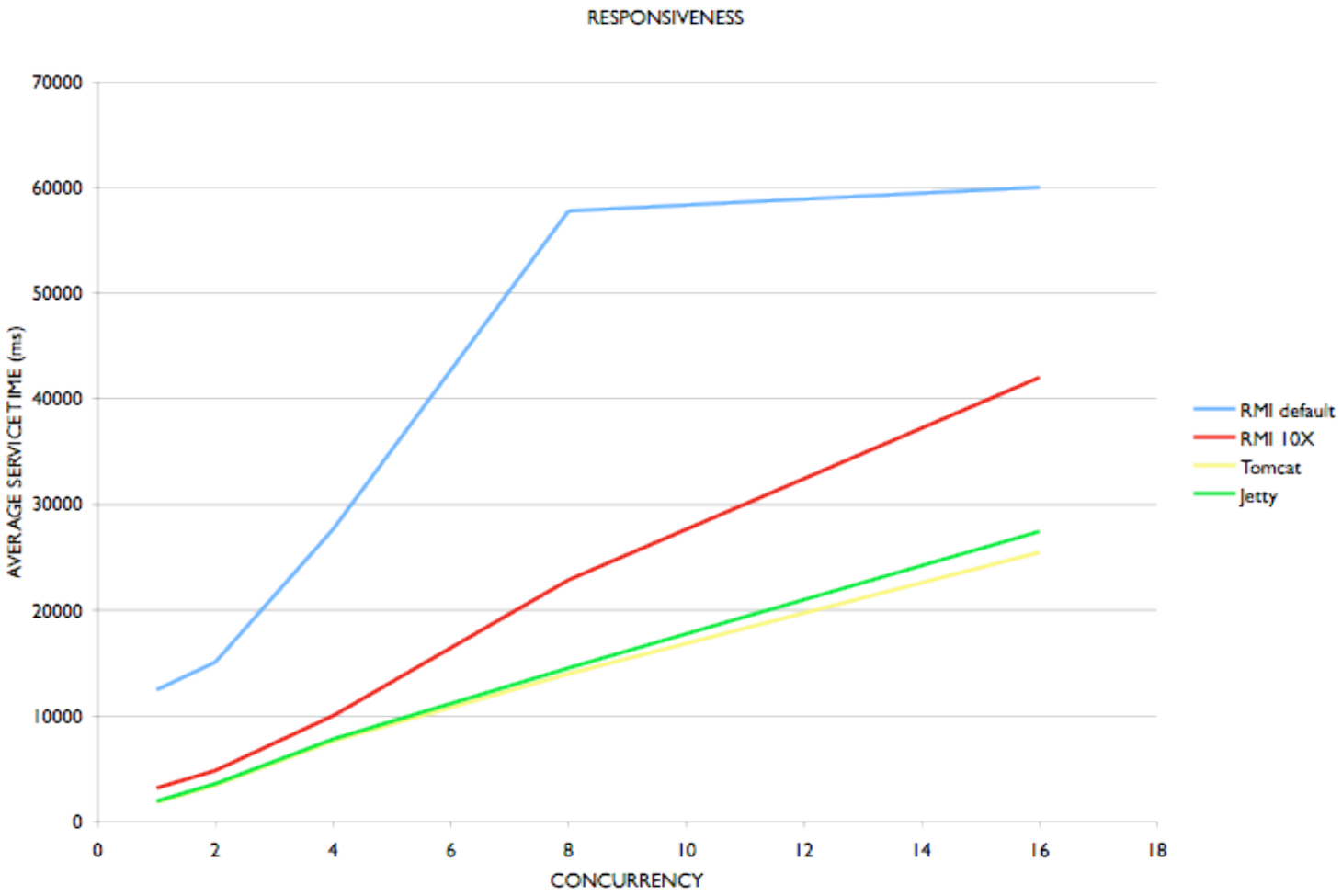
Performance

- Tested handlers: PDS File Query Handler
  - Retrieve a DIRLIST1
  - Retrieve a PDS_ZIPD of a small directory
  - Retrieve a PDS_ZIPD_SIZE
  - Extract a label
  - Retrieve a large raw file
- Tested concurrency: 1, 2, 4, 8, 16, 32, … queries at a time
- Time limit: 60 seconds for any one query to finish

Test Results

- Executive summary:
  - HTTP product servers under Tomcat totally rock the world!
  - Followed by HTTP product servers under Jetty
  - Followed by RMI with 10X block size
  - Followed by RMI with default block size


RELIABILITY
120%
100%
80%
60%
40%
20%
0%
SUCCESS RATE
0
5
10
15
20
25
30
35
CONCURRENCY
RMI default
RMI 10X
Tomcat
Jetty


RESPONSIVENESS
AVERAGE SERVICE TIME (ms)
70000
60000
50000
40000
30000
20000
10000
0
0
2
4
6
8
10
12
14
16
18
CONCURRENCY
RMI default
RMI 10X
Tomcat
Jetty

Recommendation

- HTTP product servers provide
  - Higher reliability
  - Better performance
  - Less administration
- Deploy it throughout the PDS

Summary

- Current network implementation
  - Its disadvantages
- Using HTTP instead
  - Its advantages
- Upgrade packages
- Performance